RL for Stock Trading: Q-Learning vs DQN

Chaitanya Patil
Honey Virani
Jinghua Zhu

Problem And Motivation

- **The Problem:**
- Traditional trading strategies use fixed rules
- Markets are dynamic and unpredictable
- Human traders can't monitor 24/7
- **Why RL?**
- Learns optimal actions through experience
- Adapts to changing market conditions
- Stock trading = Sequential decision making

Data Pipeline

- **Dataset:** Apple Inc. (AAPL) – Yahoo Finance
- **Period:** January 2020 – December 2024 (5 years)
- Technical Indicators Computed

Feature	Description
MA_5	5-day Moving Average
MA_20	20-day Moving Average
RSI	Relative Strength Index (14-day)
ATR	Average True Range (volatility)
Volume	Daily trading volume

Train	Validation	Test
2020 – 2022	Jan – Jun 2023	Jul 2023 – Dec 2024
~ 750 days	~ 125 days	~ 375 days

Trading Environment

- MDP Formulation

Component	Implementation
State	[Close, MA_5, MA_20, RSI, ATR, Volume, Cash/10000, Shares]
Actions	Hold (0), Buy 50%(1), Sell 50%(2)
Reward	Portfolio % change – 0.1% transaction cost

- Key Implementation Details:
  - Uses original (unnormalized) prices for actual trades
  - Transaction cost (0.1%) discourages excessive trading
  - Portfolio tracks cash + share value at each step

Baselines & Metrics

- Baseline Strategies

Stratergy	Description	Purpose
Buy & Hold	Buy at start, never sell	Industry benchmark
Random	Random action each day	Sanity check

- Evaluation Metrics

Metric	Formula	Measures
ROI	(Final – Initial) / Initial	Total return
Sharpie Ratio	Mean(r) / std(r) x $\sqrt{252}$	Risk-adjusted return
Max Drawdown	Largest peak-to-trough	Worst-case loss

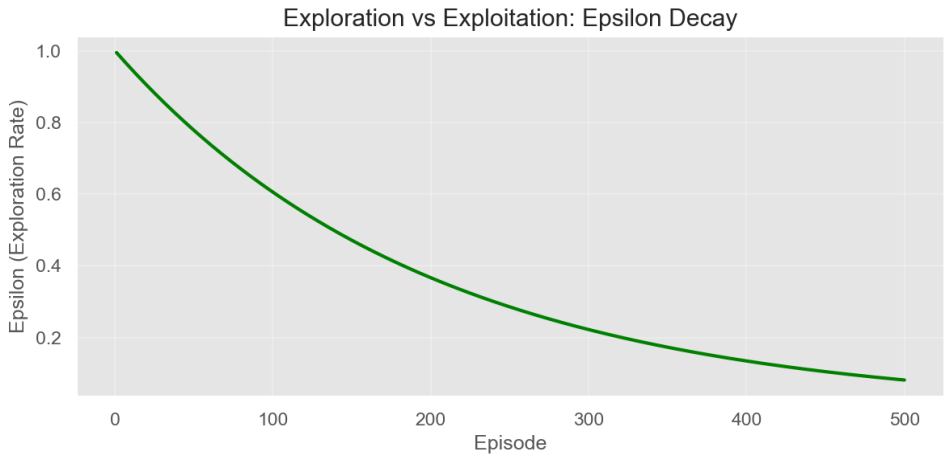
Q Learning-State Space Definition

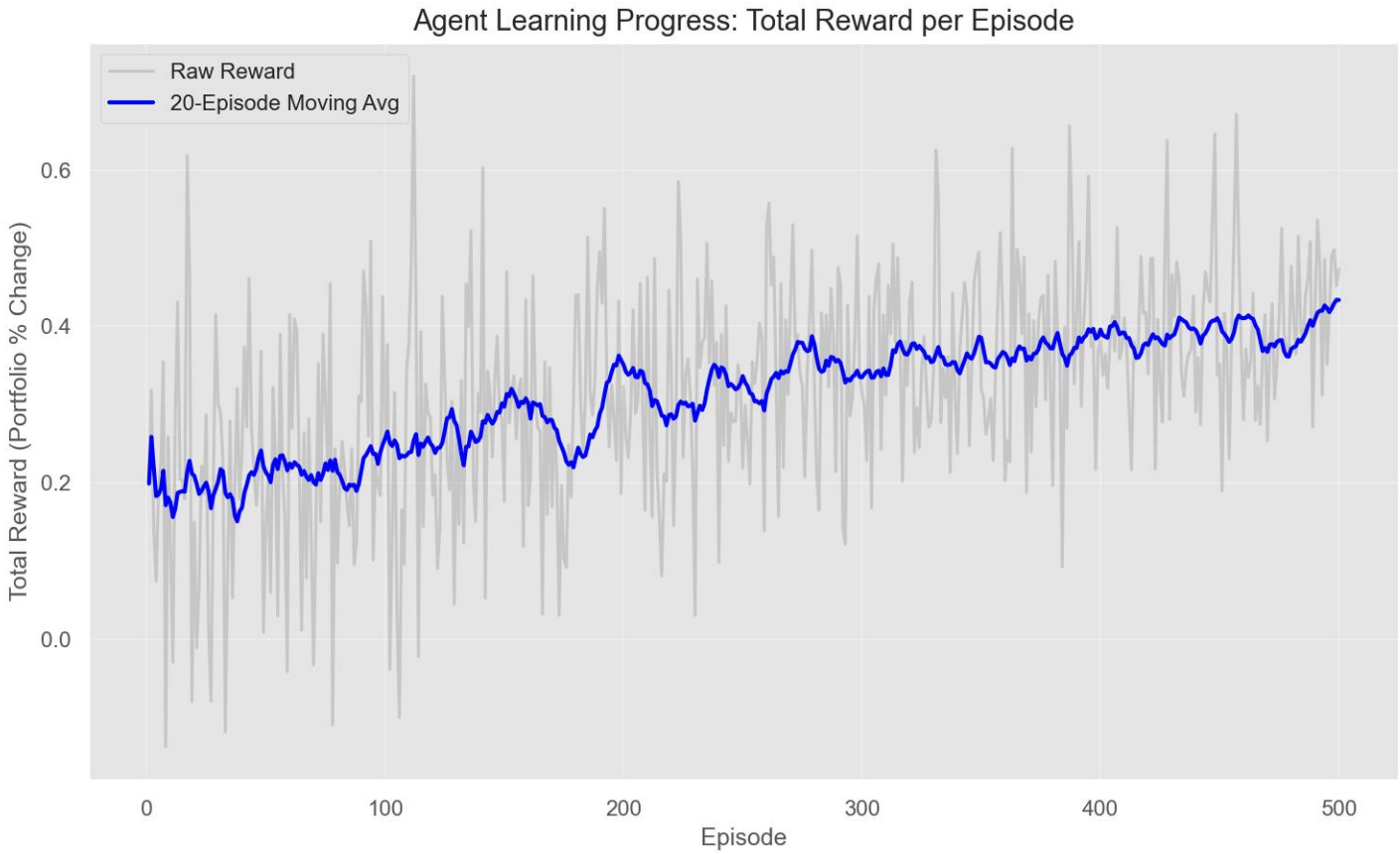
1. Price Trend:
    1. -Down (MA5 < MA20)
    2. -Flat
    3. -Up (MA5 > MA20)
2. RSI
    1. Oversold (<30)
    2. Neutral (30-70)
    3. Overbought (>70)
3. Current Position
    1. Empty
    2. Partial
    3. Full

Q Learning

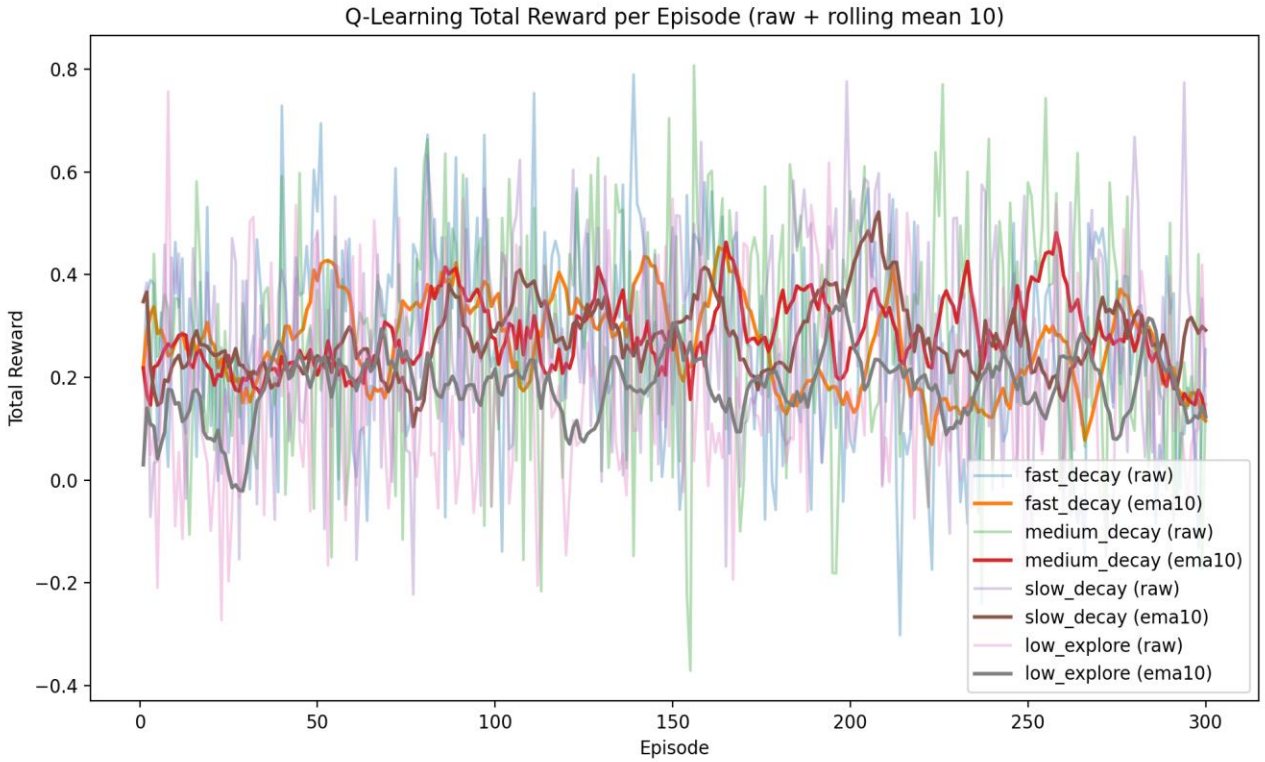
Trend: Clear upward trend in Total Reward over 500 episodes.

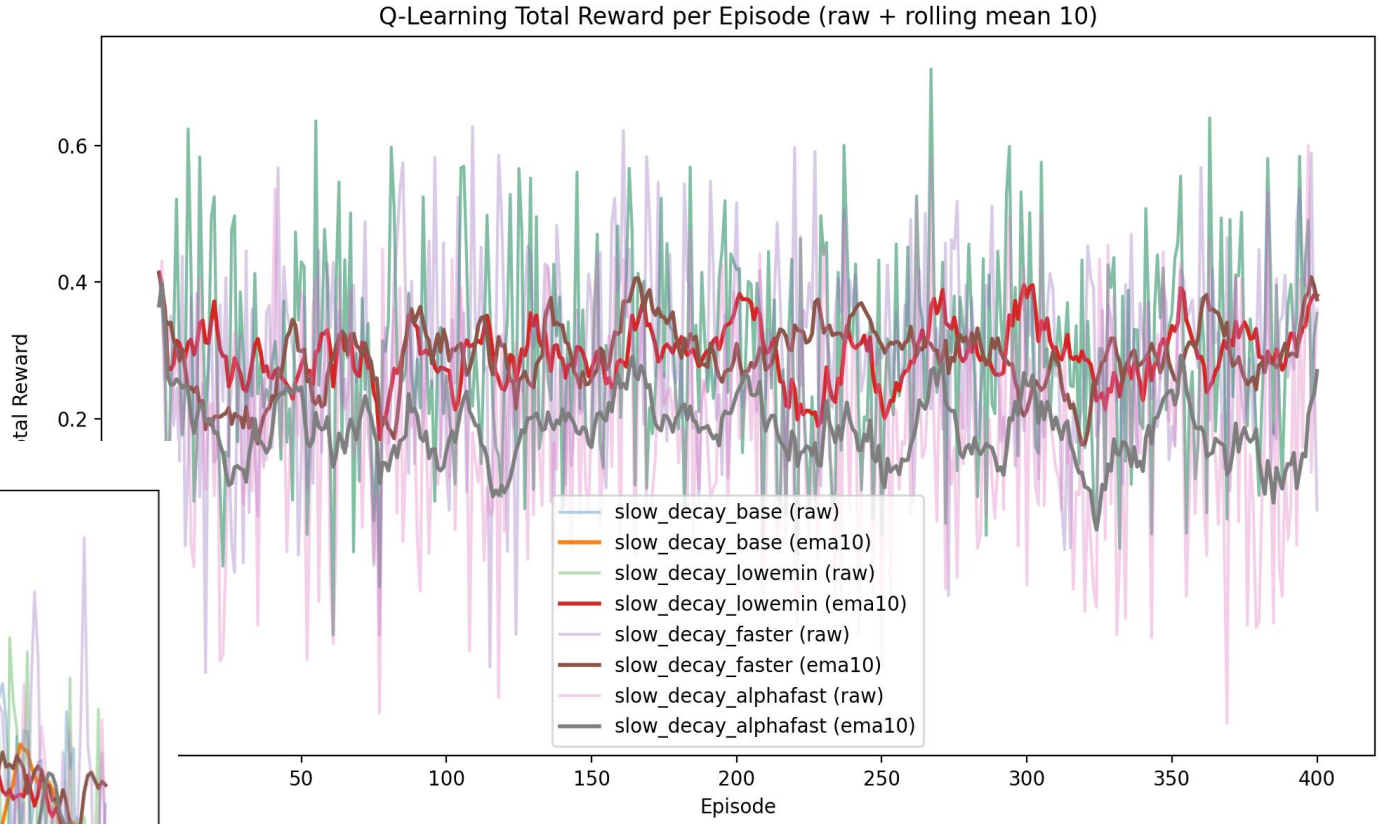
Improvement: Reward increased from ~20% to ~40%





Exploration Analysis

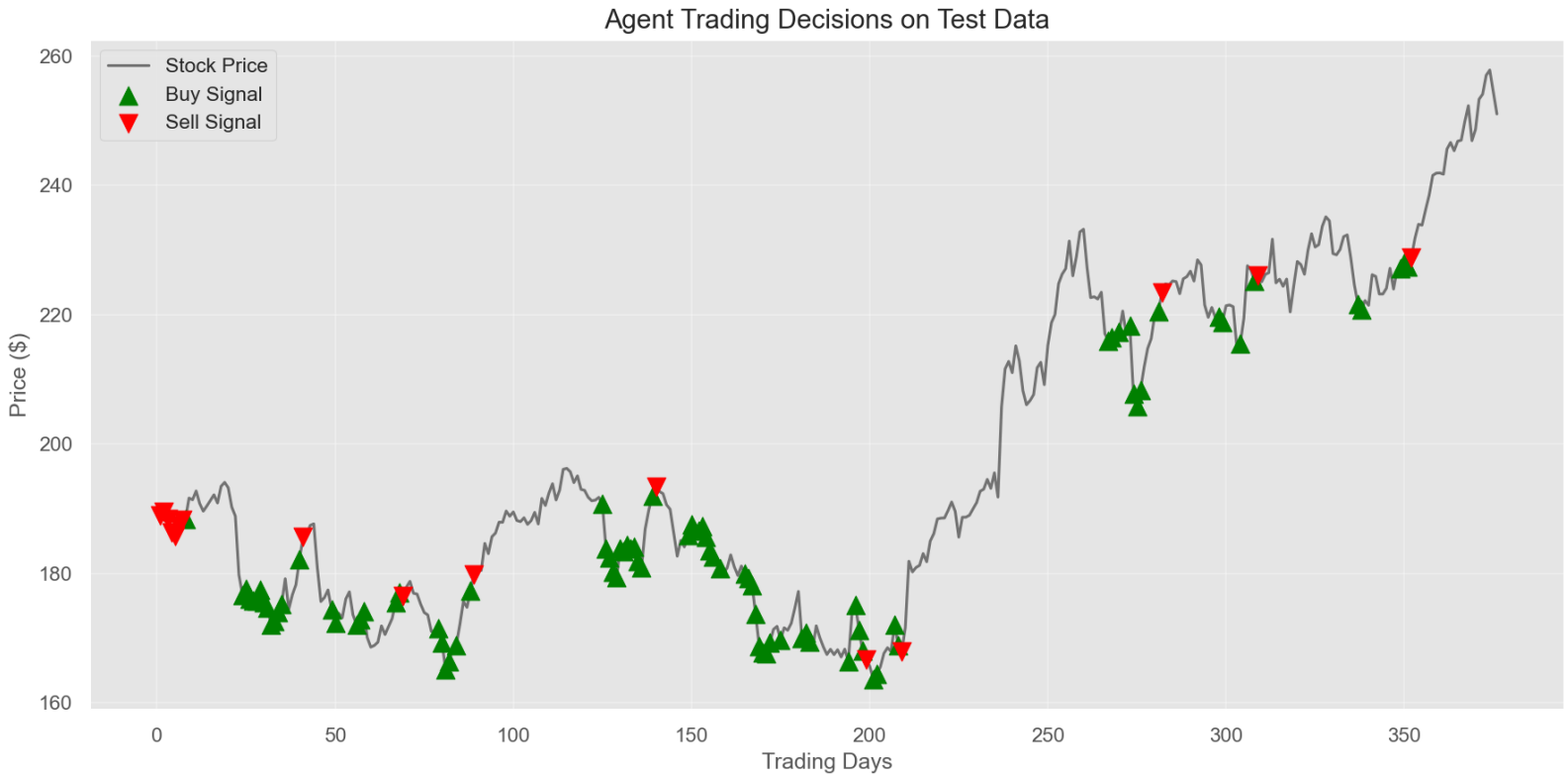




- The state and action discretization is very coarse
- The training data is fixed, and tabular Q-learning quickly learns this amount of information
- The financial market itself is highly noisy

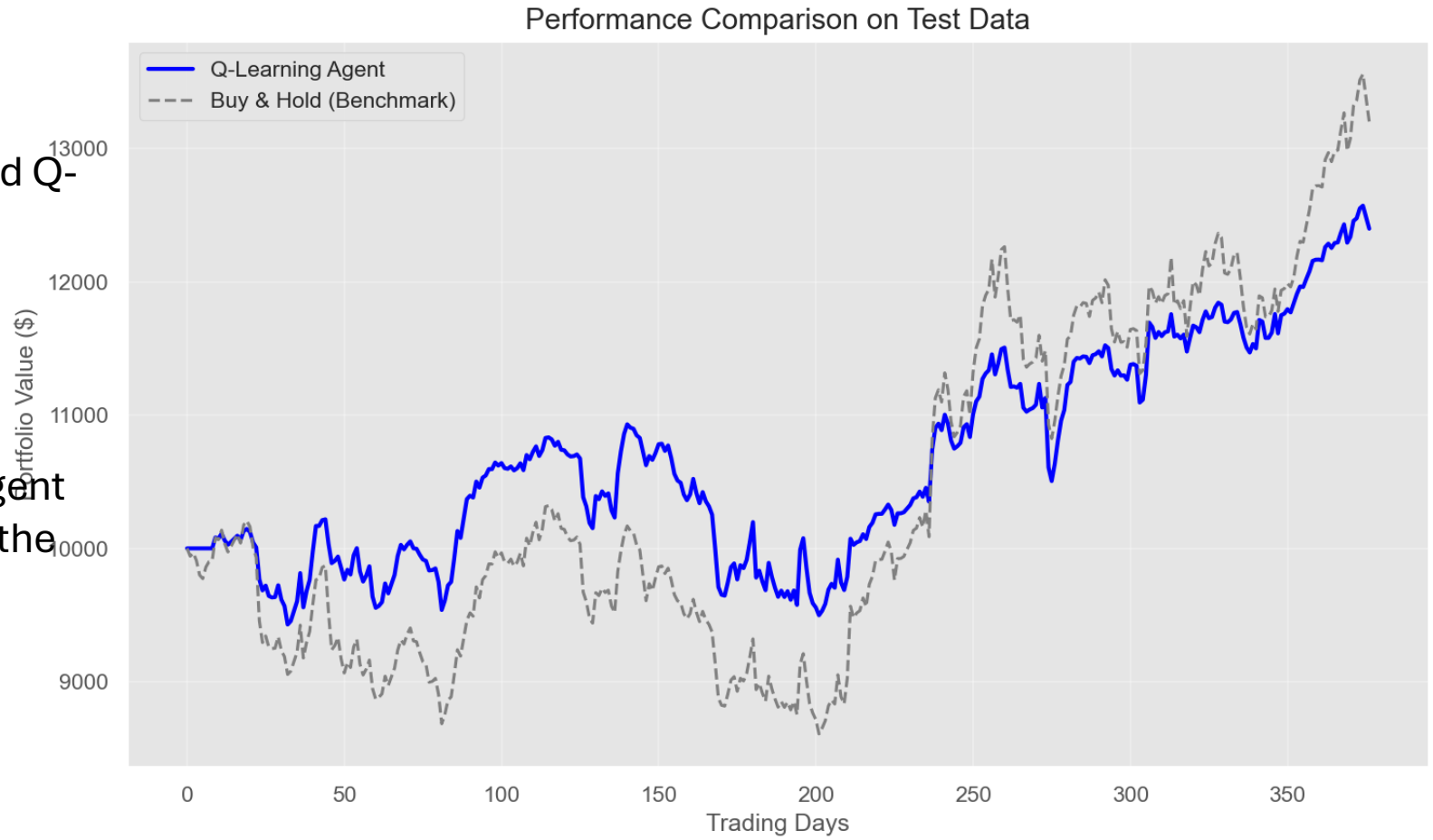
Q Learning- Trading Action Analysis

- Strength:
  - Buy low and sell high
- Weaknesses:
  - In the final strong uptrend , the agent continued to trade in and out.
  - Sell before a huge uptrend and hold until the end.



Q-Learning – Performance Evaluation

- Buy-and-Hold slightly outperformed Q-Learning
- The Agent demonstrated superior downside protection.
- While total return was lower, the Agent achieved lower volatility, validating the algorithm's ability to manage risk.



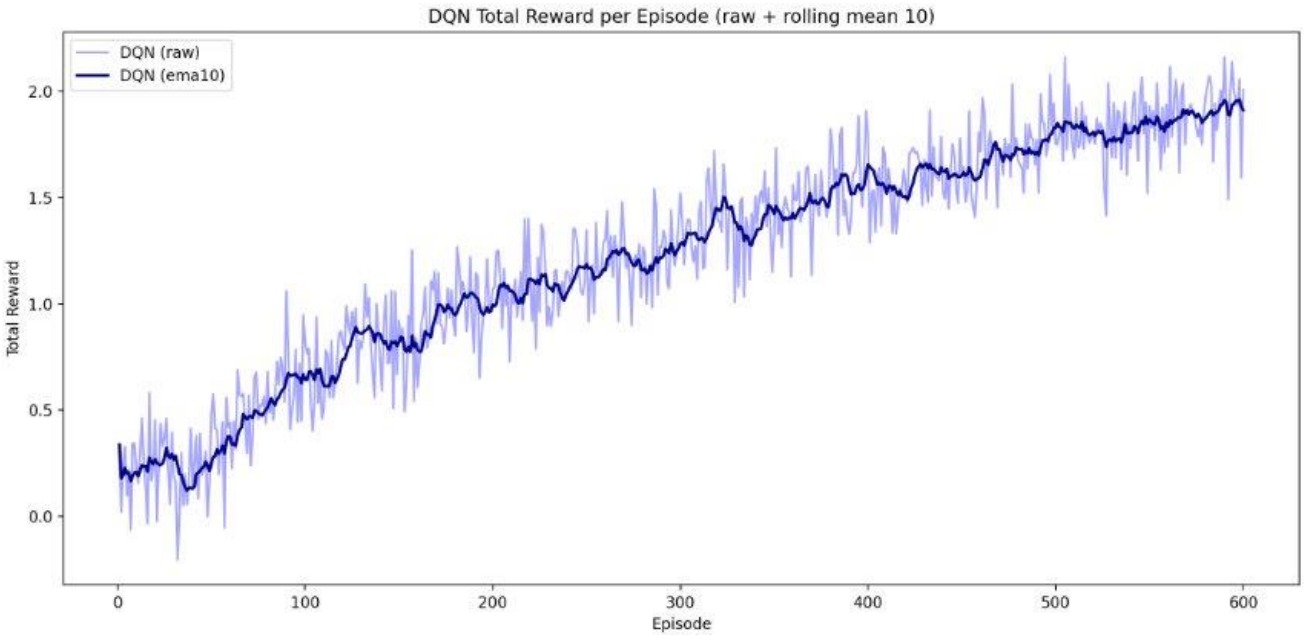
DQN

- Neural Network Architecture

Input(8) → Dense(128) → ReLU → Dense(64) → ReLU → Dense(32) → ReLU → Output(3)

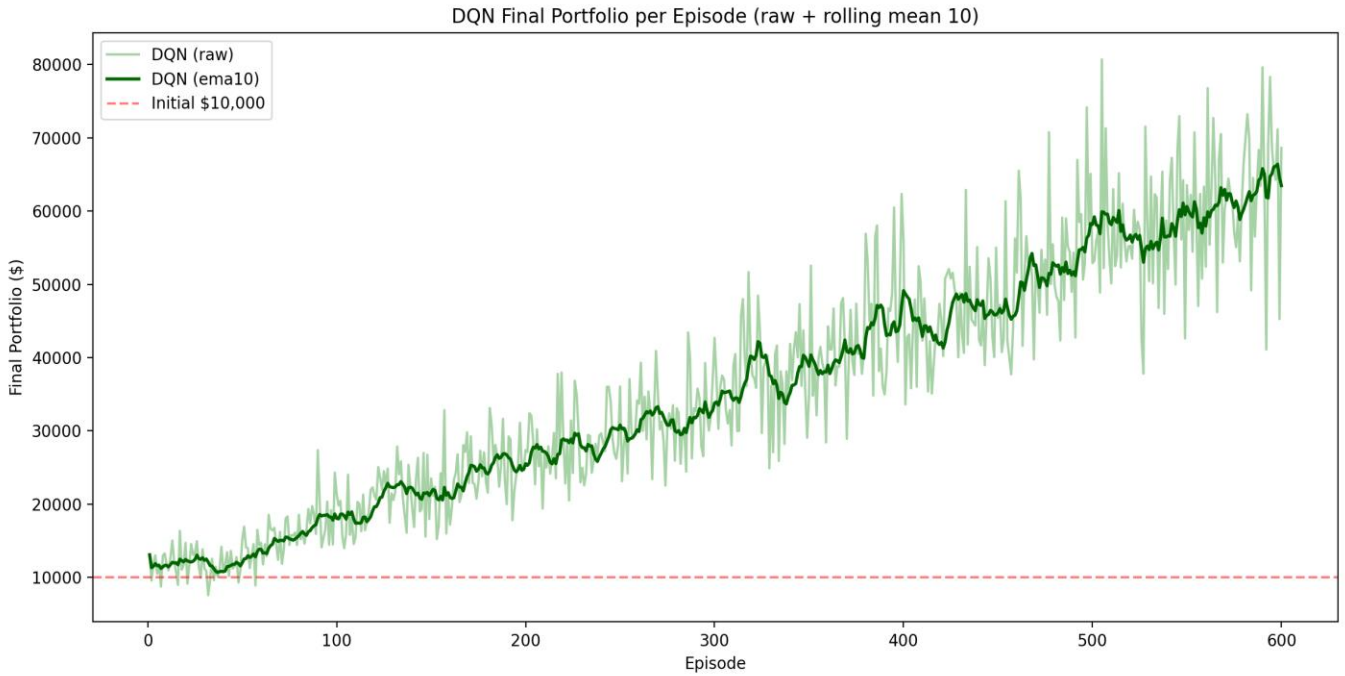
Technique	Implementation
Experience Relay	Buffer: 50000 transactions, Batch: 64
Target Network	Synced every 500 steps
Gradient Clipping	Max_norm = 1.0
Loss Function	Huber Loss (SmoothL1)

- Training Config:
  - Episodes: 600
  - Learning rate: 0.0005
  - ε: 1.0 → 0.05 (decay: 0.995)
  - Optimizer: Adam

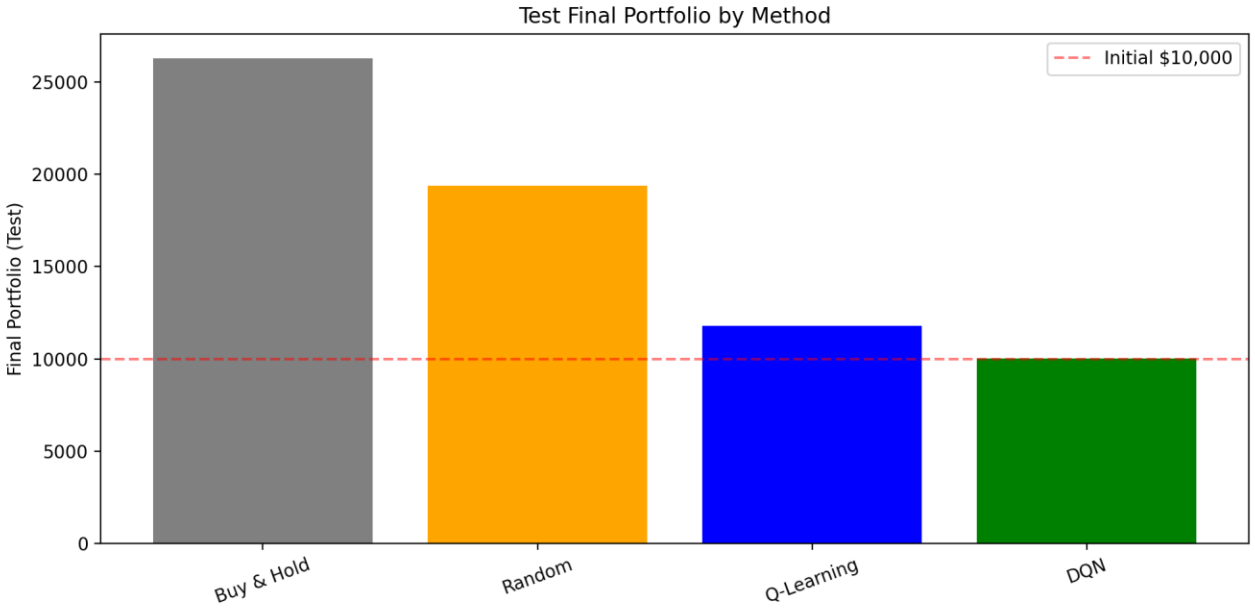


Clear Upward trend in rewards over 600 episodes
Reward increased from ~0.2 to ~1.9
Shows successful learning on training data

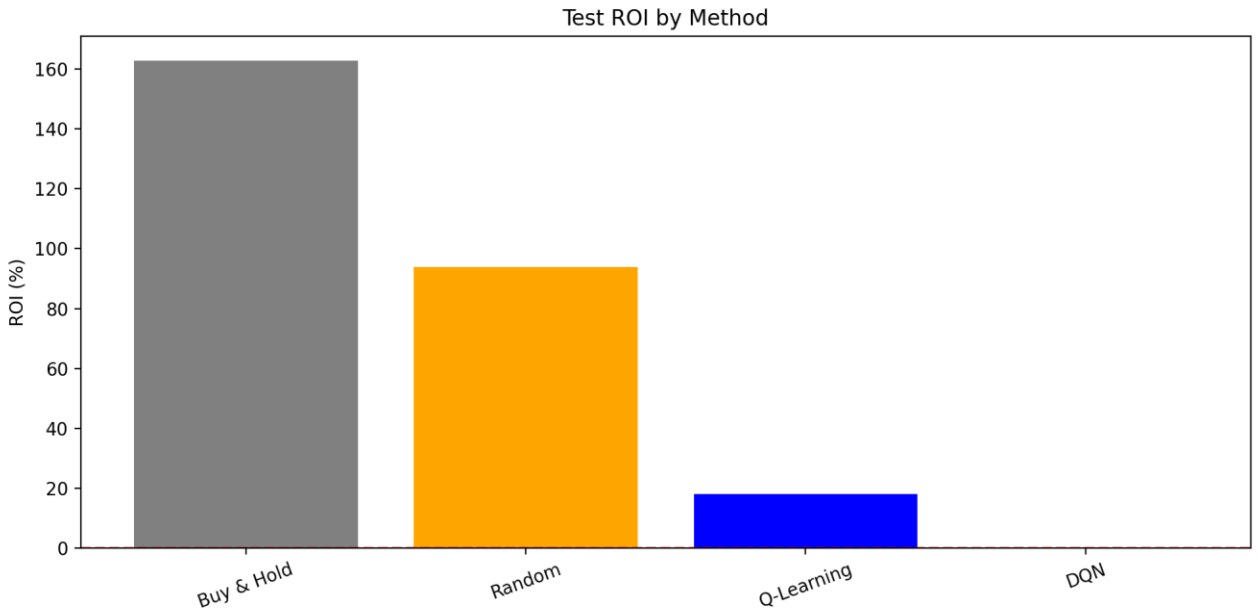
Portfolio grew from $10000 to $60000+ on training data
High variance but consistent improvement
DQN learned profitable strategies on training data



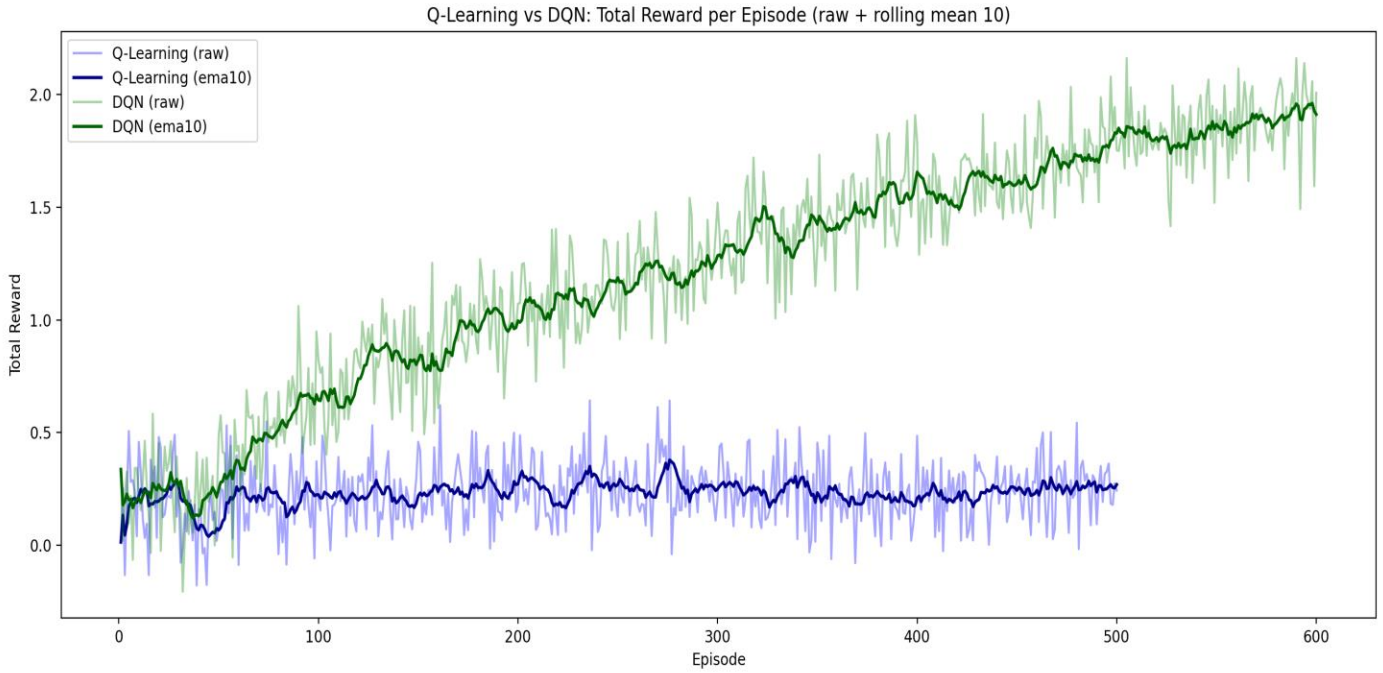
Key Results



Key Findings:
Q-learning outperformed DQN significantly (+18% vs 0%)
Both RL methods made money (above $10000 line)
Both lost to Buy & Hold in bull market



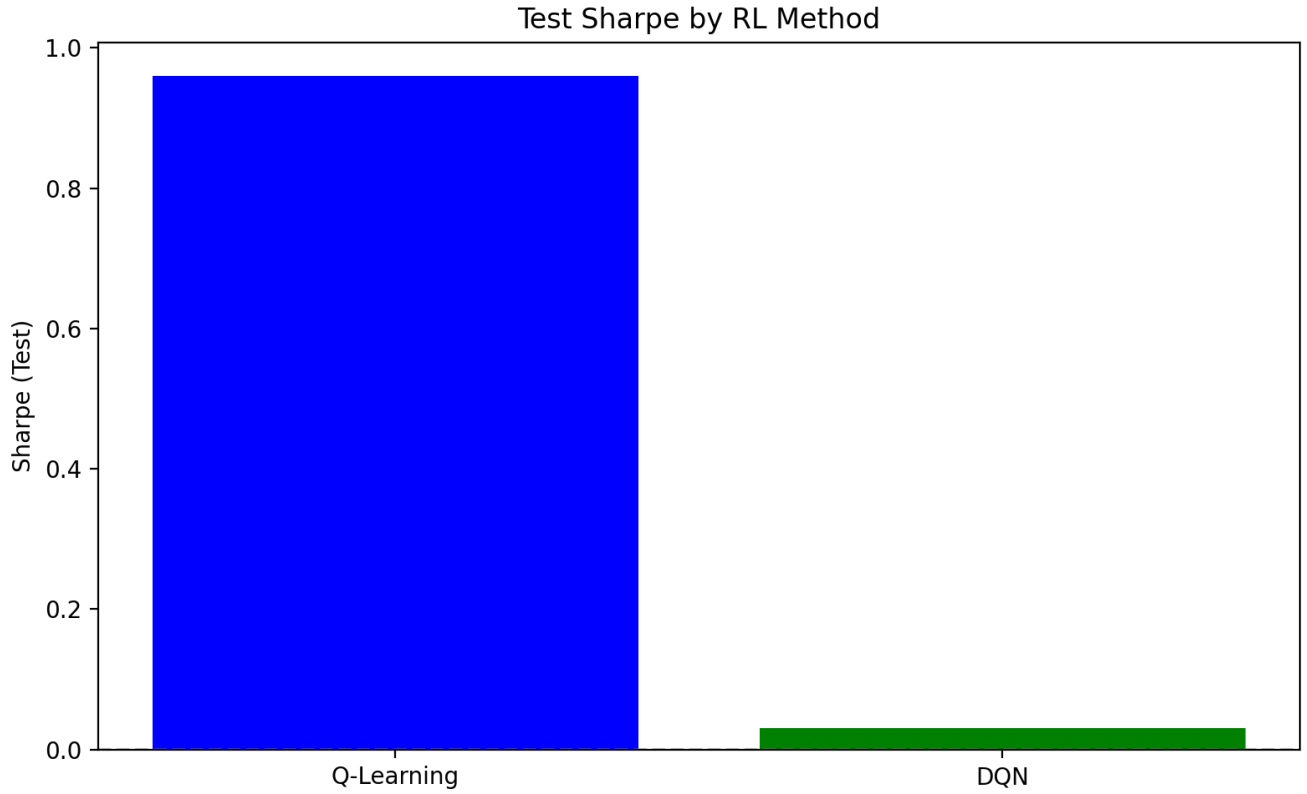
Training VS Test



During training: DQN performed much better than Q-Learning
DQN rewards: ~0.2 -> ~2.0
Q-Learning rewards: stayed flat ~0.25

Sharpe Ratio Comparison

- Q-Learning made consistent, lower-risk profits
- DQN's near-zero Sharpe means high volatility with no reward
- Even though Q-Learning's ROI (18%) seems modest, it's a solid risk-adjusted performance.



Why Buy & Hold Won

- Test Period: July 2023 – December 2024
- AAPL Performance: +163% (Massive Bull Run)
- Why RL Methods Lost to Buy & Hold:
  - Bull market optimal strategy = Buy and NEVER sell
  - Any selling = Missed gains
  - Transaction costs hurt active trading
  - RL agents learned to trade actively → suboptimal in bull markets

Conclusion

Finding	Detail
Q-Learning vs DQN	Q-Learning won (+18% vs 0%)
Training vs Test	DQN overfit, Q-Learning generalized
Complexity	Simpler discretization beat neural network

Future Work:

- Double DQN to reduce overestimation
- Regularization techniques for DQN
- Test on bear/sideways markets
- Multi-asset portfolio management

Thank You